PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-325272
(43)Date of publication of application : 22.11.2001

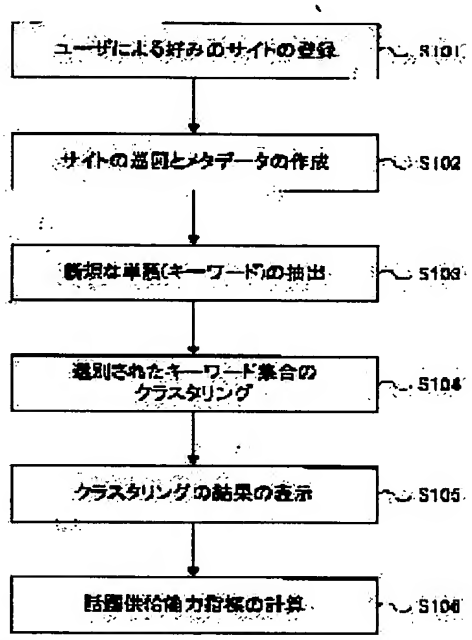
(51)Int.Cl. G06F 17/30

(21)Application number : 2000-144016
(22)Date of filing : 16.05.2000

(71)Applicant : INTERNATL BUSINESS MACH CORP <IBM>
(72)Inventor : NOMIYAMA HIROSHI

(54) INFORMATION ARRANGEMENT METHOD, INFORMATION PROCESSOR, STORAGE MEDIUM AND PROGRAM TRANSMITTER

(57)Abstract:
PROBLEM TO BE SOLVED: To freely combine plural information sources and to display information which is the topic of the conversation in an easy-to-understand form.
SOLUTION: By periodically observing the dynamically changing plural information sources acquired from the Internet, support relation between sites and the degree of the interest of an individual, etc., are taken into consideration, the more important topics of the conversation are automatically extracted from extracted information elements and they are gathered and easily understandably visualized. That is, this method is provided with a step (S102) for periodically going round the registered plural information sources and gathering the information, a step (S103) for selecting words to be the elements of the topic from the gathered information, a step (S104) for executing clustering to the set of the selected words and a step (S105) for displaying the information elements in respective clusters based on a time base on the basis of the result of clustering and displaying a main keyword from the set of the words in the respective clusters.



LEGAL STATUS

[Date of request for examination]	20.10.2000
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3606556
[Date of registration]	15.10.2004
[Number of appeal against examiner's decision	

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2001－325272
（P2001－325272A）
(43)公開日　平成13年11月22日(2001.11.22)

(51)Int.Cl.[7]	識別記号	ＦＩ	テーマコード(参考)
Ｇ０６Ｆ　17/30	２１０	Ｇ０６Ｆ　17/30	２１０Ｄ　５Ｂ０７５
			２１０Ａ
	１１０		１１０Ｆ

審査請求　有　　請求項の数19　ＯＬ　（全 14 頁）

(21)出願番号　特願2000－144016(P2000－144016)

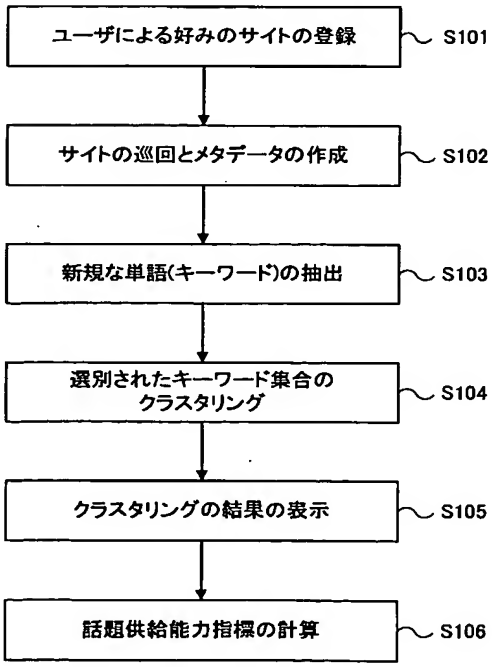
(22)出願日　平成12年５月16日(2000.5.16)

(71)出願人　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国10504、ニューヨーク州　アーモンク　（番地なし）
(74)復代理人　100104880
弁理士　古部　次郎　（外４名）

最終頁に続く

(54)【発明の名称】　情報整理方法、情報処理装置、記憶媒体、およびプログラム伝送装置

(57)【要約】
【課題】　複数の情報源を自由に組み合わせて、そこから話題となっている情報を解かり易い形で表示する。
【解決手段】　インターネットから獲得される動的に変化する複数の情報源を定期的に観察することによって、抽出される情報要素の中から、サイト間のサポート関係、個人の興味の度合いなどを考慮してより重要な話題を自動的に抽出し、それらを纏めて解かり易く視覚化する。即ち、登録された複数の情報源を定期的に巡回して情報を収集するステップ(Ｓ１０２)と、収集された情報の中から話題の要素となる単語を選別するステップ(Ｓ１０３)と、選別された単語の集合に対してクラスタリングを施すステップ(Ｓ１０４)と、クラスタリングの結果に基づいて、各クラスタにおける情報要素を時間軸に基づいて表示し、各クラスタにおける単語の集合の中から主となるキーワードを表示するステップ(Ｓ１０５)とを含む。



1

【特許請求の範囲】

【請求項１】　ネットを介して接続された情報源からの情報を整理する情報整理方法であって、
登録された複数の情報源を定期的に巡回して情報を収集する情報収集ステップと、
収集された前記情報の中から話題の要素となる単語を選別する単語選別ステップと、
選別された前記単語の集合に対してクラスタリングを施すクラスタリングステップと、
施された前記クラスタリングの結果に基づいて、各クラスタにおける情報要素を時間軸に基づいて表示すると共に、当該各クラスタにおける単語の集合の中から主となるキーワードを当該クラスタの代表キーワードとして表示する表示ステップとを含むことを特徴とする情報整理方法。

【請求項２】　前記表示ステップは、前記各クラスタにおける前記情報要素からそのテキスト部分に含まれるキーワードに基づく補足情報を表示することを特徴とする請求項１記載の情報整理方法。

【請求項３】　複数の単語が１つに縮退できる場合には縮退されたものを１つの縮退表現とする縮退ステップとを更に含み、
前記表示ステップは、前記各クラスタに新しく出現した前記縮退表現を前記補足情報として表示することを特徴とする請求項２記載の情報整理方法。

【請求項４】　前記単語選別ステップは、新しく出現した単語に対して重み付けを高くして選別することを特徴とする請求項１記載の情報整理方法。

【請求項５】　前記単語選別ステップは、特定の単語を選別した特定の情報源に対し、単語レベルで前記複数の情報源における他の情報源からのサポートを考慮して話題の要素となる単語を選別することを特徴とする請求項１記載の情報整理方法。

【請求項６】　情報を入手すべき情報源とユーザが興味のある単語とのユーザによる登録を受け付け、
登録された前記情報源に対して定期的に巡回して情報要素を入手し、
入手された前記情報要素の中からユーザの興味があるとされる単語に対して重要度を増して単語を選別し、
選別された前記単語を有する情報要素の集合に対してクラスタリングを施し、クラスタリングが施された情報要素をクラスタの結果と共に表示することを特徴とする情報整理方法。

【請求項７】　ユーザによる個々の情報源に対する興味の度合いを判断し、
判断された興味の度合いの高い情報源に出現した単語に対して重要度を増して単語を選別することを特徴とする請求項６記載の情報整理方法。

【請求項８】　情報を入手すべき複数のサイトを登録し、

2

登録された前記複数のサイトを定期的に巡回し、
内容の変化分を調べることによって巡回された複数のサイトから情報を収集し、
特定のサイトから収集された情報に対して、単語レベルで前記複数のサイトにおける他のサイトからのサポートを考慮して重要な話題を抽出することを特徴とする情報整理方法。

【請求項９】　抽出された前記重要な話題を有する情報要素に対してクラスタリングを行い、獲得された情報要素をクラスタリングの結果と共に表示することを特徴とする請求項８記載の情報整理方法。

【請求項１０】　抽出された情報要素の数に基づいて個々のサイトが提供した話題の量を計算し、
計算された話題の量に基づいて前記サイトの話題供給能力を示す指標を蓄積することを特徴とする請求項８記載の情報整理方法。

【請求項１１】　巡回すべき複数のサイトを指定する指定手段と、
前記指定手段により指定された前記複数のサイトを記憶する記憶手段と、
前記記憶手段に記憶された前記複数のサイトを定期的に巡回して情報を収集する情報収集手段と、
前記情報収集手段によって収集された情報の中から話題の要素となる単語を選別する単語選別手段と、
前記単語選別手段により選別された単語の集合に対してクラスタリングを施すクラスタリング手段と、
前記クラスタリング手段によって施されたクラスタリングの結果に基づいて、各クラスタにおける情報要素と共に、当該各クラスタにおける単語の集合の中に存在するキーワードを出力する出力手段とを含むことを特徴とする情報処理装置。

【請求項１２】　前記出力手段は、前記各クラスタにおける情報要素を時系列順に出力すると共に、当該情報要素のテキスト部分に含まれるキーワードで補足情報を出力することを特徴とする請求項１１記載の情報処理装置。

【請求項１３】　前記出力手段は、表示装置に対してまたはネットを介して接続された端末に対して出力することを特徴とする請求項１１記載の情報処理装置。

【請求項１４】　情報を入手すべき情報源とユーザが興味のある単語とのユーザによる登録を受け付ける登録受付手段と、
前記登録受付手段により受け付けられた前記情報源に対して定期的に巡回して情報要素を入手する巡回手段と、
前記巡回手段により入手された前記情報要素の中からユーザの興味があるとされる単語に対して重要度を増して単語を選別する選別手段と、
前記選別手段により選別された前記単語を有する情報要素の集合に対してクラスタリングを施すクラスタリング手段と、

前記クラスタリング手段によりクラスタリングが施された情報要素をクラスタの結果と共に表示する表示手段とを備えたことを特徴とする情報処理装置。

【請求項１５】　ユーザによる登録があった情報源またはユーザにより対応する情報要素が過去に選択された情報源に対して情報源の重要度を高く設定する設定手段とを備え、

前記選別手段は、前記設定手段によって重要度が高く設定された情報源に出現した単語に対して重要度を増して単語を選別することを特徴とする請求項１４記載の情報処理装置。

【請求項１６】　コンピュータに実行させるプログラムを当該コンピュータの入力手段が読取可能に記憶した記憶媒体において、

前記プログラムは、

登録された複数の情報源を定期的に巡回して情報を収集する処理と、収集された前記情報の中から話題の要素となる単語を選別する処理と、選別された前記単語の集合に対してクラスタリングを施す処理と、施された前記クラスタリングの結果に基づいて、各クラスタにおける情報要素を時間軸に基づいて表示すると共に、当該各クラスタにおける単語の集合の中から所定のキーワードを表示する処理とを前記コンピュータに実行させることを特徴とする記憶媒体。

【請求項１７】　前記各クラスタにおける前記情報要素からそのテキスト部分に含まれるキーワードに基づく補足情報を当該各クラスタに新しく出現した縮退表現を用いて表示する処理とを含むことを特徴とする請求項１６記載の記憶媒体。

【請求項１８】　コンピュータに実行させるプログラムを当該コンピュータの入力手段が読取可能に記憶した記憶媒体において、

前記プログラムは、

情報を入手すべき複数のサイトを登録する処理と、登録された前記複数のサイトを定期的に巡回する処理と、内容の変化分を調べることによって巡回された複数のサイトから情報を収集する処理と、収集された情報に対して、単語レベルで他のサイトからのサポートを考慮して重要な話題を抽出する処理とを前記コンピュータに実行させることを特徴とする記憶媒体。

【請求項１９】　コンピュータに実行させるプログラムを記憶する記憶手段と、当該記憶手段に記憶された当該プログラムを送信する送信手段とを備えたプログラム伝送装置であって、

前記記憶手段に格納される前記プログラムは、登録された複数の情報源を定期的に巡回して情報を収集する処理と、収集された前記情報の中から話題の要素となる単語を選別する処理と、選別された前記単語の集合に対してクラスタリングを施す処理と、施された前記クラスタリングの結果に基づいて、各クラスタにおける情報要素を

時間軸に基づいて表示すると共に、当該各クラスタにおける単語の集合の中から所定のキーワードを表示する処理とを備え、前記送信手段によって送信可能に構成されることを特徴とするプログラム伝送装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、情報源からの情報検索に係り、特に、インターネット上の複数の情報源から話題となっている情報を抽出して視覚化する方法等に関する。

【０００２】

【従来の技術】近年、インターネットの整備に伴い、ユーザが入手できる情報の量は膨大となっている。この膨大な情報源の中から、ユーザの欲する情報を出来るだけ早く、正確に、そしてユーザの使い易い形で整理して提供する情報検索技術は、ますます重要性が増している。

【０００３】従来の情報検索技術として、例えば、登録された情報源(サイト)から情報を伝える要素(リンクとそのタイトル、テキストの連続等)を抽出し、そのテキスト部分を言語解析するものがある。また、検索サービス、ニュースなどの情報提供サービスを行なうポータルサイト(portal site)を利用して話題を抽出する技術も存在する。このポータルサイトでは、人手による作成によって話題となっているキーワードを提供するサービスを行なっており、例えば検索者にとっての話題であるキーワードランキング等を利用してユーザに提供するサービスが存在する。

【０００４】また、文献１(J. Kleinberg. Authoritative sources in a hyperlinked environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, . Also appears as IBM Research Report RJ 10076, May 1997.)には、インターネット上のある一瞬における静的な構造上の参照関係(サポート)を考慮して重要度の計算を行なう技術について開示されている。ここでは、指定された検索式に対する権威のあるページ(Authority)と、権威のあるページを多く含むページ(Hub)を抽出している。また、文献２(柳瀬,仲尾 "メールマガジンを利用した注目ニュースの自動抽出," 57-20, p. 151- 158,情報処理学会情報学基礎研究会予稿集, 3/22/2000.)には、複数の情報源(メールマガジン)を情報源とし、そこから注目ニュースの自動抽出を行なう技術について開示されている。ここでは、クラスタリングされた結果の重要度として情報源の数が多い(メールマガジンの種類が多い)という指標が用いられている。

【０００５】一方、特開平８－２８７０７４号公報では、継続的に発行される文書等、最近の文書に現われる未登録語の発生頻度をリアルタイムに監視し、現在注目を集めつつあるトピックに関係する用語および文書を利用者に定義する技術について開示されている。また、特開平１１－１４３８９２号公報では、文章中に出現する

キーワードの重みとカテゴリ情報を考慮した重みを合成してキーワードの重みを生成する技術について示されている。更に、特開平１１-１４３７９６号公報では、メーリングリストサービスにおいて、各メーリングリストでやり取りされている主な話題を抽出する技術が開示されている。

【０００６】

【発明が解決しようとする課題】このように、情報を整理して話題となっていることを自動的に抽出し、それらを解かり易く表示することは非常に有用であり、従来から幾つかの提案がなされている。しかしながら、上述したポータルサイトなどでは、重要な分野のニュースに関して話題の抽出を人手で行なっているが、単一のサイトだけでは情報の評価基準が偏っている可能性があり、重要な情報を見逃す恐れがあったり、１つの話題に関する全ての情報が得られない恐れがある。この恐れを回避するために複数のサイトを見ようとすると、情報が重複してしまう問題がある。また、観点がまちまちとなることから、ユーザに対して理解を容易にするためには別の観点から整理し直す必要性がある。更に、多くの読者が期待できない分野のニュースに関しては、人手で情報を整理するというようなサービスは行なわれておらず、ユーザが自分自身で複数のサイトを集め、纏めることが必要となってしまう。

【０００７】一方、上記文献１では、話題になっているものを抽出する技術については含まれておらず、また、参照関係の重み付けに検索式中のキーワードを利用しているが、結果に単語そのものを含めるものではない。文献２では、単語が新しく出現したかどうかは考慮していない。また、情報源の数が多いという指標をクラスタの重要度の判定に用いており、単語の重要度の判定ではないので、サポートの導入はクラスタリングの結果に影響を与えることができない。

【０００８】更に、特開平８-２８７０７４号公報では、あくまでも静的な情報源(文書)を対象としており、インターネットなどの同一の情報源からの情報は別の文書として認識されてしまう。このため、重要度の計算においては文書の数のみを考慮しているに過ぎない。また、文書間の構造化を含まず、ある文書集合に頻出する単語を識別するなどの文書分類の性質を利用することができない。更に、辞書に登録されていない単語あるいは複合語だけを抽出の対象としており、登録後を組み合わせて文章となっている場合は新出概念として抽出できない。その結果、例えば、「日本ＩＢＭが新しいデータベース製品を発表」などの文は全て登録語で構成されており、この手法では新出概念として抽出することができないのである。また更に、関連する新出語があってもその類似性が考慮されていないため、関連する新語を同時に見ることができず、関連する新語の登録作業が別々になり、その都度同じような文書集合が提示され、作業の効

率化が図れない。また、特開平１１-１４３８９２号公報では、時間的な側面への考慮がなく、動的な情報源に対する考慮もない。更に、特開平１１-１４３７９６号公報では、対象がメーリングリストに限定されており、また、単一のメーリングリストを対象とするものに過ぎず、複数の情報源から話題を抽出するような情報検索は困難である。

【０００９】本発明は、以上のような技術的課題を解決するためになされたものであって、その目的とするところは、複数の情報源を自由に組み合わせて、そこから話題となっている情報を解かり易い形で表示することにある。また他の目的は、ユーザの興味に沿ったクラスタリングの結果を得ることにある。

【００１０】

【課題を解決するための手段】かかる目的のもと、本発明は、インターネットなどから獲得される動的に変化する複数の情報源(ＵＲＬなどで参照される)を定期的に観察することによって、抽出される情報要素の中から、サイト間のサポート関係、個人の興味の度合いなどを考慮してより重要な話題を自動的に抽出し、それらを纏めて解かり易く視覚化するものである。即ち、本発明は、ネットを介して接続された情報源からの情報を整理する情報整理方法であって、登録された複数の情報源を定期的に巡回して情報を収集する情報収集ステップと、収集された情報の中から話題の要素となる単語を選別する単語選別ステップと、選別された単語の集合に対してクラスタリングを施すクラスタリングステップと、施されたクラスタリングの結果に基づいて、各クラスタにおける情報要素を時間軸に基づいて表示すると共に、各クラスタにおける単語の集合の中から主となるキーワードをクラスタの代表キーワードとして表示する表示ステップとを含むことを特徴としている。

【００１１】この表示ステップは、各クラスタにおける情報要素からそのテキスト部分に含まれるキーワードに基づく補足情報を表示することを特徴としている。また、複数の単語が１つに縮退できる場合には縮退されたものを１つの縮退表現とする縮退ステップとを更に含み、表示ステップは、各クラスタに新しく出現した縮退表現を補足情報として表示することを特徴としている。これらの発明によれば、得られた情報をユーザに対してより解かり易く視覚化して表現することが可能となる点で好ましい。

【００１２】また、単語選別ステップは、新しく出現した単語に対して重み付けを高くして選別することを特徴とすれば、新出したニュースをユーザに対して優先的に提供できる。更に、この単語選別ステップは、特定の単語を選別した特定の情報源に対し、単語レベルで複数の情報源における他の情報源からのサポートを考慮して話題の要素となる単語を選別することを特徴とすれば、話題となっている情報を選別してユーザに提供できる点で

優れている。

【００１３】本発明を他の観点から把えると、本発明が適用された情報整理方法は、情報を入手すべき情報源とユーザが興味のある単語とのユーザによる登録を受け付け、登録された情報源に対して定期的に巡回して情報要素を入手し、入手された情報要素の中からユーザの興味があるとされる単語に対して重要度を増して単語を選別し、選別された単語を有する情報要素の集合に対してクラスタリングを施し、クラスタリングが施された情報要素をクラスタの結果と共に表示することを特徴とすることができる。更に、ユーザによる個々の情報源に対する興味の度合いを判断し、判断された興味の度合いの高い情報源に出現した単語に対して重要度を増して単語を選別することを特徴とすることができる。このユーザによる興味の度合いの判断としては、例えば、ユーザによる特定サイトの指定の他、例えば、ユーザによって対応する情報要素が過去において選択されたサイトを興味の度合いが高いとして判断すること等が可能である。

【００１４】更に他の観点から把えると、本発明が適用された情報整理方法は、情報を入手すべき複数のサイトを登録し、登録された複数のサイトを定期的に巡回し、例えば指定された期間にて新出した単語等による内容の変化分を調べることによって巡回された複数のサイトから情報を収集し、特定のサイトから収集された情報に対して、単語レベルで複数のサイトにおける他のサイトからのサポートを考慮して重要な話題を抽出することを特徴とすることができる。また、抽出された重要な話題を有する情報要素に対してクラスタリングを行い、獲得された情報要素をクラスタリングの結果と共に表示することを特徴とすることができる。このクラスタリングの結果の表示とは、例えば、各クラスタ毎に時系列表示するものや、各クラスタの代表キーワードや補足情報を表示すること等が挙げられる。また、抽出された情報要素の数に基づいて個々のサイトが提供した話題の量を計算し、計算された話題の量に基づいてサイトの話題供給能力を示す指標を蓄積することを特徴とすれば、例えば蓄積された話題供給能力に基づいてサイトや単語の重み付けを行なうこと等に利用できる点で好ましい。更に、この応用としては、サイトを話題供給能力指標順に並べ、また、その数値を表示すること等が可能である。

【００１５】一方、上記目的を達成するために、本発明が適用される情報処理装置は、巡回すべき複数のサイトを指定する指定手段と、指定された複数のサイトを記憶する記憶手段と、記憶された複数のサイトを定期的に巡回して情報を収集する情報収集手段と、収集された情報の中から話題の要素となる単語を選別する単語選別手段と、選別された単語の集合に対してクラスタリングを施すクラスタリング手段と、施されたクラスタリングの結果に基づいて、各クラスタにおける情報要素と共に、各クラスタにおける単語の集合の中に存在するキーワードを出力する出力手段とを含むことを特徴とすることができる。

【００１６】ここで、この出力手段は、各クラスタにおける情報要素を時系列順に出力すると共に、情報要素のテキスト部分に含まれるキーワードで補足情報を出力することを特徴とすれば、抽出された個々の話題がどのように変化していったかを解かり易く出力することができる点で優れている。尚、この出力手段は、表示装置に対して表示する態様の他、ネットを介して接続された端末に対して電子情報として出力する態様とすることが可能である。

【００１７】他の観点から把えると、本発明が適用された情報処理装置は、情報を入手すべき情報源とユーザが興味のある単語とのユーザによる登録を受け付ける登録受付手段と、受け付けられた情報源に対して定期的に巡回して情報要素を入手する巡回手段と、入手された情報要素の中からユーザの興味があるとされる単語に対して重要度を増して単語を選別する選別手段と、選別された単語を有する情報要素の集合に対してクラスタリングを施すクラスタリング手段と、クラスタリングが施された情報要素をクラスタの結果と共に表示する表示手段とを備えたことを特徴とすることができる。また、ユーザによる登録があった情報源またはユーザにより対応する情報要素が過去に選択された情報源に対して情報源の重要度を高く設定する設定手段とを備え、選別手段は、この設定手段によって重要度が高く設定された情報源に出現した単語に対して重要度を増して単語を選別することを特徴とすることができる。

【００１８】一方、本発明は、コンピュータに実行させるプログラムをコンピュータの入力手段(例えばＣＤ-ＲＯＭドライバ等)が読取可能に記憶した記憶媒体(例えばＣＤ-ＲＯＭ等)において、このプログラムは、登録された複数の情報源を定期的に巡回して情報を収集する処理と、収集された情報の中から話題の要素となる単語を選別する処理と、選別された単語の集合に対してクラスタリングを施す処理と、施されたクラスタリングの結果に基づいて、各クラスタにおける情報要素を時間軸に基づいて表示すると共に、所定のキーワードとして、例えば各クラスタにおける単語の集合の中から主となるキーワードをクラスタの代表キーワードとして表示する処理とをコンピュータに実行させることを特徴としている。ここで、この各クラスタにおける情報要素からそのテキスト部分に含まれるキーワードに基づく補足情報を各クラスタに新しく出現した縮退表現を用いて表示する処理とを含むことを特徴とすれば、ユーザに対して更に解かり易い表示を提供することができる点で好ましい。

【００１９】また本発明は、コンピュータに実行させるプログラムをコンピュータの入力手段が読取可能に記憶した記憶媒体において、このプログラムは、情報を入手すべき複数のサイトを登録する処理と、登録された複数

のサイトを定期的に巡回する処理と、内容の変化分を調べることによって巡回された複数のサイトから情報を収集する処理と、収集された情報に対して、単語レベルで他のサイトからのサポートを考慮して重要な話題を抽出する処理とをコンピュータに実行させることを特徴とすることができる。

【００２０】更に、本発明は、コンピュータに実行させるプログラムを記憶する記憶手段と、この記憶手段に記憶されたプログラムをインターネット等を介してユーザ端末に対して送信する送信手段とを備えたプログラム伝送装置であって、この記憶手段に格納されるプログラムは、登録された複数の情報源を定期的に巡回して情報を収集する処理と、収集された情報の中から話題の要素となる単語を選別する処理と、選別された単語の集合に対してクラスタリングを施す処理と、施されたクラスタリングの結果に基づいて、各クラスタにおける情報要素を時間軸に基づいて表示すると共に、各クラスタにおける単語の集合の中から所定のキーワードを表示する処理とを備え、この送信手段によって送信可能に構成されることを特徴とすることができる。

【００２１】

【発明の実施の形態】以下、添付図面に示す実施の形態に基づいてこの発明を詳細に説明する。まず、本実施の形態におけるシステム構成の説明に入る前に、本方式の理解を容易にするために、その概要について説明する。図１は、本実施の形態における情報抽出/表示手法の概要を示す図である。本手法は、個人が自由に情報源を選択し、更に、それらに対し興味の度合いに応じて自由に重要度を付与することによって情報を整理し、自分専用の情報サイト(Personal Portal)あるいは、特定の分野専用のサイト(Vertial Portal)を自動的に実現するものである。そのために、まず、ユーザによって好みのサイトの登録がなされる(ステップ１０１)。登録する際には、例えば、その名前とその参照(ＵＲＬ：Uniform Resource Locators)を指定する。次にシステムは、登録されたサイトを指定された時刻に定期的に巡回し、その内容をデータベースに登録されているものと比較する。その内容が異なった場合には、新しいバージョンとして登録し、メタデータを作成する(ステップ１０２)。このメタデータは、ＵＲＬで参照される内容から、情報を選ぶ要素を抽出したものである。

【００２２】次に、登録されているサイトの集合中における個々のサイトにおいて、指定された期間の直前のバージョンと、指定された期間のバージョンに出現したキーワードをカウントし、キーワードの集合に重み付けを施して新規な単語(キーワード)を抽出する(ステップ１０３)。その後、選別されたキーワードの集合を、個々のキーワードが含まれる情報要素集合の包含関係および付与された重みを用いてクラスタリングを行う(ステップ１０４)。このクラスタリングとは、何らかの観点で

意味のある集合で分けていく作業と言える。そして、このクラスタリングの結果から、各クラスタのキーワード集合の主となる代表キーワード(ホットワード)を表示し、情報要素集合を時間順に表示すると共に、補足情報としてのキーワード(サブワード)を用いてクラスタリングの結果を表示する(ステップ１０５)。この一連の処理によって、サイト間のサポート関係、個人の興味の度合いなどを考慮してより重要な話題を自動的に抽出でき、また、それらを纏めて解かり易く視覚化することが可能となる。その後、このようにして抽出されたクラスタに対して、そのキーワードの重要度に基づいて、個々のサイトがどれくらい話題を提供する能力があるかを示す指標である話題供給能力指標を計算する(ステップ１０６)。これにより、話題抽出の際に計算された重要度を用いて、話題供給能力の高いサイト、あるいは特定の単語に対してより話題供給能力の高いサイトを提示することができる。

【００２３】次に、システム構成を用いて、本手法を更に詳述する。図２は、本実施の形態におけるシステムの全体構成を説明するための図である。本システムは、インターネット１０に接続されるパーソナルコンピュータ(ＰＣ)等にてアプリケーションソフトの処理プログラムとして実行される。また、インターネット１０に接続されたユーザのＰＣ端末に情報を提供するサーバとして構成することも可能である。この処理プログラムによる出力は、ユーザのＰＣ端末ではディスプレイに表示される場合の他、サーバである場合にはインターネット１０を介してユーザのＰＣ端末に提供するように構成される。尚、本実施の形態では、ユーザのＰＣ端末における処理の流れを中心に説明している。更に、このシステムを実行する処理プログラムは、ハードディスクドライブ(図示せず)に格納され、実行時にはメインメモリ(図示せず)にロードされてＣＰＵ(図示せず)によって処理されるのが一般的である。また、この処理プログラムは、例えばＣＤ-ＲＯＭ(図示せず)による記憶媒体を介してユーザのＰＣ端末等に供給される場合の他、例えばインターネット１０を介してユーザが処理プログラムをダウンロードすることによって提供される形態も考えられる。

【００２４】図２において、符号１１はユーザが登録したサイトを保存する登録サイトＤＢ(データベース)、１２は前述したメタデータを格納するメタデータＤＢ、１３はキーワードの重要度から計算により得られたサイトの重要度を格納するサイト話題供給能力ＤＢ、１７はユーザが指定した好みのキーワードあるいはサイトの重要度を格納するユーザ指定重み付けＤＢであり、これらは、例えばＰＣに設けられたハードディスクドライブ等の記憶手段の一部を利用している。１４は登録されたサイトをインターネット１０から自動巡回するクローラである。１５は登録されたサイトのメタデータを保存、管理するバージョン管理機能付きＤＢＭＳ(データベース

マネージメントシステム)であり、ＨＴＭＬ(Hypertext Markup Language)の中から情報要素を抽出し、そのテキスト部分を解析して、それに含まれるキーワードとその分類を保存するメタデータ作成機構２０を備えている。１６はメタデータＤＢ１２中に蓄積されているデータへのアクセス手段を提供するメタデータアクセスメソッドである。また、３０は新規情報抽出表示機構であり、メタデータＤＢ１２に蓄積された情報を元に、新しい話題を抽出して表示する機構である。

【００２５】登録サイトＤＢ１１に登録されるサイトは、前述したようにユーザの好みによって登録されるサイトである。ユーザは、登録する際にその名前とその参照(ＵＲＬ)を指定する。図３は、登録されたサイトの例を示している。図３に示される例では、４つのサイトが登録されており、その登録の形式はＸＭＬ(eXtensible Markup Language)である。尚、例えば、特定のポータルサイトのディレクトリ・リストをカットアンドペーストして登録する方法がユーザにとっては簡単な操作と言えるであろう。

【００２６】クローラ１４では、登録サイトＤＢ１１に登録されたサイトを指定された時刻に定期的に巡回する。例えば、毎日午前７時３０分に巡回する等である。指定されたサイト全体を同一時刻で巡回してもよいし、個々のサイトに対して異なった時刻を指定することも可能である。バージョン管理機能付きＤＢＭＳ１５は、クローラ１４による巡回時に内容が異なっていた場合に新しいバージョンとして管理し、更にメタデータ作成機構２０によってそれに対するメタデータを作成して、その結果をメタデータＤＢ１２に保存している。このように、サイトの新しいバージョンが作成された場合には、そのメタデータが作成される。このメタデータは、前述したようにＵＲＬで参照される内容から情報を運ぶ要素を抽出したものである。それには、リンクとそのテキスト部分、あるいは連続したテキスト部分がある。これら情報要素のテキスト部分に関しては属性抽出が適用され、キーワードとその分類が抽出される。

【００２７】図４は、メタデータ作成機構２０の構成を更に詳述したものである。このメタデータ作成機構２０は、図４に示されるように、ＨＴＭＬなどの入力ファイルからメタデータを作成して出力ファイルとして出力している。符号２１は情報要素抽出機構であり、ＨＴＭＬなどの内容を解析して情報要素となるもの(リンク、テキストなど)を抽出している。２２は属性抽出機構であり、情報要素抽出機構２１により抽出された情報要素のテキストからキーワードを抽出し、それにカテゴリを付与している。この属性抽出機構２２は、形態素解析機構２３、キーワード抽出機構２４、およびキーワード分類機構２５を備えている。この形態素解析機構２３は情報要素抽出機構２１により抽出された情報要素のテキスト部分を単語に分割している。キーワード抽出機構２４は形態素解析機構２３により分割された結果の単語列からキーワードとなるものだけを抽出している。キーワード分類機構２５はキーワード抽出機構２４により抽出されたキーワードの分類を付与する機能を備えている。

【００２８】図５は作成されたメタデータの例としてリンクの例を示した図である。また、図６は作成されたメタデータの例としてテキストブロックの例を示している。図５において、リンクの場合におけるＨＴＭＬファイル中の表現は、リンク先を示すタグを用いた a タグで示されており、抽出された情報要素は anchor タグによって構成される。また、図６において、テキストブロックの場合におけるＨＴＭＬファイル中の表現は、テキスト表現であり、抽出された情報要素は text タグによって構成されている。以上の処理によって、登録サイトＤＢ１１に登録されたサイトにおいて、クローラ１４による巡回時に変化があった場合には、その全ての内容と、メタデータ作成機構２０によってそこから作成されたメタデータがメタデータＤＢ１２に登録される。また、内容に変更のあった日時(ウェブサーバから更新日時が得られる場合にはその日時、得られない場合には巡回した日時など)が、同様にメタデータＤＢ１２に保存される。

【００２９】次に、新規情報抽出表示機構３０にて新規な単語の抽出とそのクラスタリングが行なわれる。図７は、この新規情報抽出表示機構３０における構成を説明するための図である。同図において、符号３１はキーワード統計機構であり、メタデータＤＢ１２から得られる、指定されたサイトに対するメタデータから、指定された期間内のバージョンに新たに出現した情報要素に含まれるキーワードと、指定された期間の直前のバージョンに含まれる情報要素中に含まれるキーワードとをカウントしている。情報要素が新たに出現したかどうかの判断は、リンクに対しては、異なったＵＲＬのリンクが出現したか、あるいは、同じＵＲＬがすでに存在していたがその対応するテキストが異なった場合に新しいリンクと判断される。テキストブロックに対しては、異なったテキストが出現したかどうかによって判断される。３２はキーワード重要度計算機構であり、抽出されたキーワードに対して重要度を付与している。このキーワード重要度計算機構３２では、サイト話題供給能力ＤＢ１３を参照して、サイトの重要度を加味した重要度の設定を行なうことが可能である。３３はクラスタリング機構であり、抽出された重要度付きキーワードを用いてクラスタリングを行なっている。この抽出されたクラスタに対して、後述するようにキーワードの重要度に基づいて重要度を計算して、その結果をサイト話題供給能力ＤＢ１３に格納している。３４はクラスタリング結果表示機構であり、クラスタリングの結果を表示する機能を有する。

【００３０】図８は、指定された期間とバージョンとの関係を示した図である。図７に示したキーワード統計機

構３１では、登録サイトＤＢ１１に登録されているサイト集合中の個々のサイトにおいて、図８に示す指定された期間の直前のバージョンと指定された期間のバージョンとに出現したキーワードがカウントされる。ここでは、指定された開始日時に直前のバージョン(Version N-3)に含まれるカウント(Fs(w))と、その後のバージョン(Version N-2からVersion N)に含まれるカウント(Fn(w))が区別される。キーワード重要度計算機構３２では、これらのキーワード集合に重み付けを施すことによって、新規なキーワードであるかどうかの判断が行なわれる。選別する方法は、例えば、単語の重要度やサイトの重要度といった重みを単独あるいは組み合わせて、それが閾値以下のものを排除する方法が考えられる。

【００３１】単語の重要度としては、以下のような検討例が考えられる。

（ａ） 単純な新出語の割合(Fn(w)/(Fs(w)+ Fn(w)))を考慮する。

（ｂ） 過去のバージョン(Version N-3以前の全てのバージョン)におけるキーワードの情報量を計算し、情報量が低いキーワードは重要度を下げる。これにより、例えば「新製品の発売情報」等における“発売”等、必ず個々の情報に付与されるような単語は、重要度を低くすることができる。

（ｃ） 単語が複数のサイトに含まれるか(複数のサイトからサポートされているか)どうかを考慮する。

（ｄ） ユーザ指定による重み付けを行なう。即ち、ユーザが特に興味のある(或いは興味のない)単語を重要度と共に登録し、それが出現した場合は重要度を高く(低く)する。

指定の方法は、ユーザが明示的に個々のサイトに対して重要度を記述する方法、或いは最終的に表示されたクラスタリングの結果を表示したときに、その対応する情報要素が選択された場合は、その情報要素を含むサイトの重みを高くする方法等が考えられる。

【００３２】サイトの重要度としては、ユーザによる個々のサイトに対する重要視の度合いを基準とする方法がある。例えば、ユーザが特に興味のある(或いは興味のない)サイトを登録し、そのサイトに出現した単語は重要度を高くする(低くする)等である。指定の方法は、ユーザが明示的に個々のサイトに対して重要度を記述する方法、或いは、最終的に表示されたクラスタリングの結果を表示したときに、その対応する情報要素が選択された場合には、その情報要素を含むサイトの重みを高くする方法がある。

【００３３】次に、選別されたキーワード集合のクラスタリングについて説明する。図７に示したクラスタリング機構３３では、キーワード統計機構３１で選別されたキーワード集合を、キーワード重要度計算機構３２で付与された重みを用いてクラスタリングが行なわれる。このクラスタリングの手法としてはどのようなものでも構わないが、クラスタリングの前処理として、複数のキーワードが全く同一のキーワード集合を含み、かつ、それらのキーワードが１つに縮退できる場合には、縮退されたものを１つのキーワードとしている。

【００３４】ここで、縮退とは、例えば以下のようなものを含むものである。

- 正書

正書辞書を用いて正書に変換する。

"コンピュータ", "コンピューター"などの表記の揺れ → 正書"コンピュータ"に変換する。

- 同義語

同義語辞書を用いて正規表現に変換する。

"米国", "アメリカ合衆国" → "米国"

- 複合語

すべてのテキスト中で複合語として隣接して出現する単語を１つの複合語に変換する。

"小渕", "首相" → "小渕首相"

- 依存構造

すべてのテキスト中で同じ依存関係を持つ語を１つの表現に変換する。ケースマーカ(case marker)が得られる場合はそれも付与する。ケースマーカは、日本語の場合は助詞など、英語の場合は前置詞などが対応する。以下の例ではケースマーカとして助詞“が”が付与されている。

"内閣", "総辞職" → "内閣が総辞職"

【００３５】次に、クラスタリングの一例を説明する。ここでは、まず最初に選別されたキーワードを重要度順にソートする。そして、個々のキーワードに対して、そのキーワードが含まれる情報要素を割り当てる。その後、包含関係(強い包含関係と弱い包含関係)の決定がなされる。この包含関係を決定する際には、個々のキーワードは必ず重要度の高いキーワードに含まれるということを前提としている。この包含関係の決定では、全てのキーワードについて、それより重要度が高い全てのキーワードに対して包含関係の有無を調べる。包含関係の有無は、キーワードに対応付けられた情報要素を集合として見た場合、その共通する要素の割合が閾値より大きい場合に強い包含関係があるものとする。また、何らかの共通の情報要素はあるがその割合が閾値に満たないものは、弱い包含関係にあるものとする。強い包含関係があるとみなされたキーワードは１つのクラスタに纏められる。弱い包含関係にあるとみなされたキーワードは別のクラスタとなる。ここで、弱い包含関係に含まれる情報要素集合は、より重要度の高いキーワードに対応付けられたクラスタの情報要素集合に含まれるものは除かれている。また、そのキーワードは、より重要度の高いクラスタのキーワード集合に加えられる。

【００３６】図９は、このようなクラスタリングの結果として得られたものの構造の例と、その解釈について示している。図９に示す例では、キーワード１は、キーワ

ード２およびキーワード３と強い包含関係がある。また、キーワード４とキーワードN－１とも強い包含関係がある。また、キーワード４はキーワード３と弱い包含関係にある。クラスタリングの結果により、クラスタ１、クラスタ２およびクラスタmの集合が形成されている。このクラスタ１のキーワード集合としては、強い包含関係のあるキーワード１～３が纏められ、また、補足として弱い包含関係のあるキーワード４も集合化されている。一方、情報要素集合では、強い包含関係のあるキーワード１～３に対応する情報要素集合１～３で集合化されており、情報要素集合４は除かれている。この情報要素集合４はフル状態でテキストが出力されることから、情報量を減らす意味で弱い包含関係にある情報要素集合が除かれている。

【００３７】図１０は、クラスタリングが行なわれた具体例を示した図である。ここでは、クラスタ１～３の３つのクラスタを示しており、それぞれ、キーワード集合と情報要素集合が形成されている。このクラスタ２とクラスタ３は、クラスタ１に対して弱い包含関係にある。

【００３８】次に、クラスタリングの結果の表示について説明する。図７に示したクラスタリング結果表示機構３４は、上述したクラスタリングの結果から、各クラスタのキーワード集合の主となるキーワード(最も重要度の高いキーワード)を、クラスタの代表キーワード(ホットワード)として表示する。更に、そのクラスタに含まれる情報要素集合の中から、情報集合を時間順に表示する。その際、その情報要素のテキスト部分に含まれるキーワードで補足情報をサブワードとして表示する。この補足情報は、そのクラスタのキーワード集合に含まれるキーワードの、単一縮退表現、或いは複数のキーワード或いは縮退表現、が最初に出現した場合に表示する。キーワードおよび縮退表現の表示順序は、テキストに現われた出現順と同じ順序とする。

【００３９】図１０で挙げた具体例で説明すると、クラスタ１の表示において、まず最も古い情報要素から表示される。その情報要素「開発ツール、e-コマース、オペレーティング・システム、データベース、ロータス製品、ネットワーク関連」には、キーワード集合の中の１つのキーワードである「データベース」しか含まれていないので、サブワードは表示されない。次の情報要素「JDBCコンプライアント・リレーショナル・データベース管理システム(DB2, Oracle等)に格納されたリレーショナル・テーブルのセットにより、XMLアクセスサービスLightweight Extractor(XLE)は、データベースよりデータを抽出し、その抽出データをXML文書に変換、アセンブルします。」には、キーワード集合の中の「データベース」と「ＤＢ」が含まれている。ここには複数のキーワードが含まれているので、これらを用いてサブワードが作成される。その表示の順番は、情報要素集合の中におけるテキスト中の順番で表示されることから、「ＤＢ，データベース」となる。もしも、テキスト中にこれらのキーワードが連続して出現している場合には、その縮退表現「ＤＢデータベース」(カンマがない)で表示される。このサブワードは記憶され、クラスタ１の表示において「データベース」、「ＤＢ」のみが含まれる場合には、再び表示されることはない。

【００４０】次に、そのクラスタと弱い包含関係のあるクラスタがあれば、それを表示する。ホットワードの表示においては、包含関係があることを示すために「段付け」を行なう。サブワードの表示も同様にして行なわれる。このようにして、全てのクラスタが表示される。弱い包含関係のクラスタと弱い包含関係にあるクラスタのホットワードとは、そのレベルの数だけ「段付け」されて表示される。

【００４１】図１１は、これらの一連の処理によって得られた表示例を示した図である。図１１に示す表示例では、最も左端にホットワード５１が表示され、その隣にサブワード５２が表示される。また、日付５３から理解できるように、最も古い情報要素から表示されている。また、参照記事５４では、情報要素としてテキストブロックとアンダーラインで示されるリンク文章とが表示されている。更に、ホットワードの１段目のキーワード「データベース」に対して包含関係のある「バージョン」と「ＤＢ」とは、「段付け」されて１段下げて表示されているのが理解できる。このように、本実施の形態では、クラスタリングされた結果を時系列に表示し、各クラスタの主となるキーワード(ホットワード)に加えて、新しく出現した縮退表現を補足情報(サブワード)として表示し、対応する情報要素を時系列順に表示している。これによって、ユーザに対してより新しく、且つ、ユーザの欲する情報を、整理された状態で提供することが可能となる。

【００４２】最後に、本実施の形態では、話題供給能力指標の計算を行なっている。即ち、このようにして抽出されたクラスタに対して、そのキーワードの重要度に基づいて重要度を計算することができる。この結果得られた重要度は、サイト話題供給能力ＤＢ１３に対して加算的に蓄積され、更新が行なわれ、サイトの重要度の計算に用いることが可能となる。その際、過去の値を減少させることによってできるだけ最新の状況を反映するようにする。より具体的には、抽出されたクラスタに含まれる単語、情報要素の数、或いはその重みを組み合わせることによって個々のサイトが提供した話題の量を計算し、それを元にサイトの話題供給能力を示す指標として蓄積している。また、クラスタに含まれる単語に関しても、個々のサイトにおける単語別の話題供給能力指標として蓄積する。また、サイトを個々のサイトに付与された話題供給能力指標順に並べる、或いはその数値を表示することによって、サイトがどれくらい新しい情報を提供してきたかをユーザに提示する。更に、個々のサイト

17

に付与された単語毎の話題供給能力指標を用いて、個々のサイトにおける特定の単語に対する情報供給能力指標を提示することも可能である。また更に、個々のサイトに付与された話題供給能力指標付き単語集合に対し、ユーザが指定したキーワードに一致するサイトを表示することによって、ユーザが求めるキーワードに対して話題供給能力の高いサイトを提示することも可能である。

【００４３】このように、本実施の形態によれば、複数の情報源を自由に組み合わせて、そこから話題となっている情報を取り出すことで、単一情報源ではなく情報源の集合内で話題となっている情報を獲得することができる。即ち、複数のサイトを登録し、それを定期的に巡回し、その内容の変化分を調べることによって、より重要な話題を抽出することができる。また、複数の情報源のサポートを考慮することで単語に対する重みが変わるので、クラスタリングの結果が変わり、サイト集合内でより一般的なクラスタを得ることが可能となる。即ち、単語レベルで他のサイトからのサポートを考慮することにより、より重要な話題を抽出することができる。同様に、単語やサイトに対するユーザの興味の度合いを変えることによって、ユーザの興味に沿ったクラスタリングの結果を得ることができる。更に、例えば、補足情報を用いて獲得されたテキストをクラスタリングの結果と共に表示することによって、抽出された個々の話題がどのように変化していったかを解かり易く表示することが可能となる。

【００４４】

【発明の効果】以上説明したように、本発明によれば、複数の情報源を自由に組み合わせて、そこから話題となっている情報を解かり易い形で表示することが可能となる。

【図面の簡単な説明】

【図１】　本実施の形態における情報抽出/表示手法の概要を示す図である。

【図２】　本実施の形態におけるシステムの全体構成を説明するための図である。

【図３】　登録されたサイトの例を示した図である。

【図４】　メタデータ作成機構２０の構成を更に詳述した図である。

【図５】　作成されたメタデータの例としてリンクの例を示した図である。

【図６】　作成されたメタデータの例としてテキストブロックの例を示した図である。

【図７】　新規情報抽出表示機構３０における構成を説明するための図である。

【図８】　指定された期間とバージョンとの関係を示した図である。

【図９】　クラスタリングの結果として得られたものの構造の例とその解釈について示した図である。

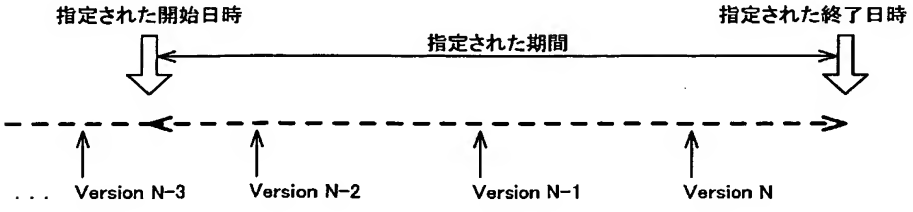
【図１０】　クラスタリングが行なわれた具体例を示した図である。

【図１１】　これらの一連の処理によって得られた表示例を示した図である。

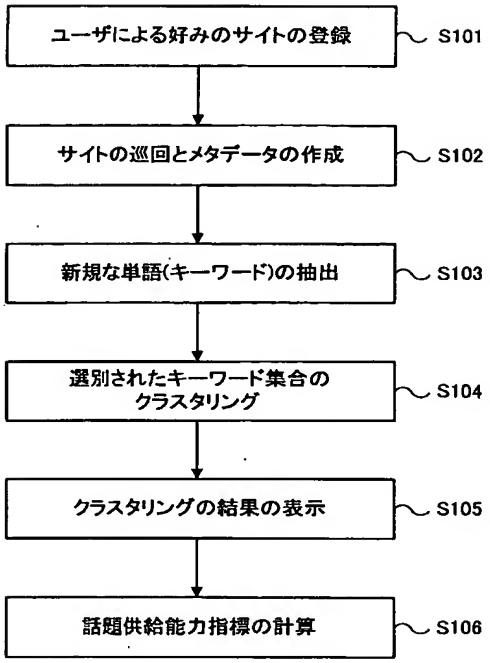
【符号の説明】

１０…インターネット、１１…登録サイトＤＢ、１２…メタデータＤＢ、１３…サイト話題供給能力ＤＢ、１４…クローラ、１５…バージョン管理機能付きＤＢＭＳ、１６…メタデータアクセスメソッド、１７…ユーザ指定重み付けＤＢ、２０…メタデータ作成機構、２１…情報要素抽出機構、２２…属性抽出機構、２３…形態素解析機構、２４…キーワード抽出機構、２５…キーワード分類機構、３０…新規情報抽出表示機構、３１…キーワード統計機構、３２…キーワード重要度計算機構、３３…クラスタリング機構、３４…クラスタリング結果表示機構、５１…ホットワード、５２…サブワード、５３…日付、５４…参照記事

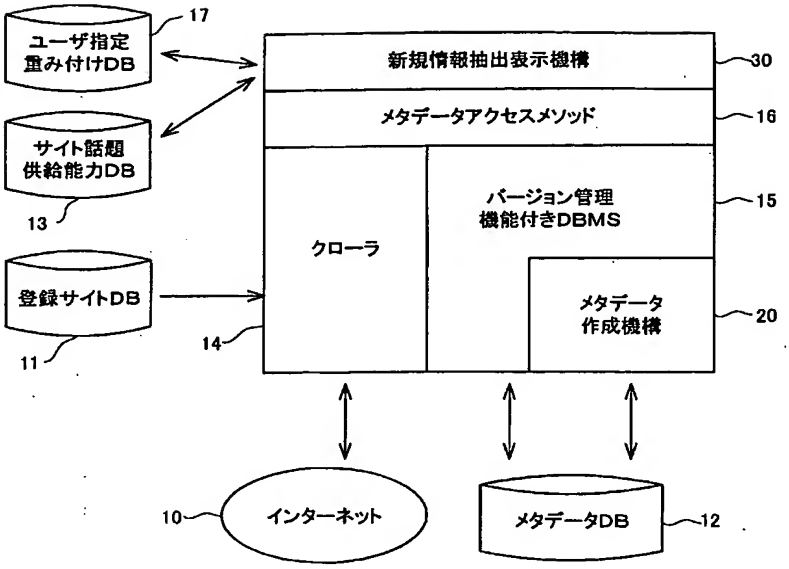
【図８】



【図１】



【図２】

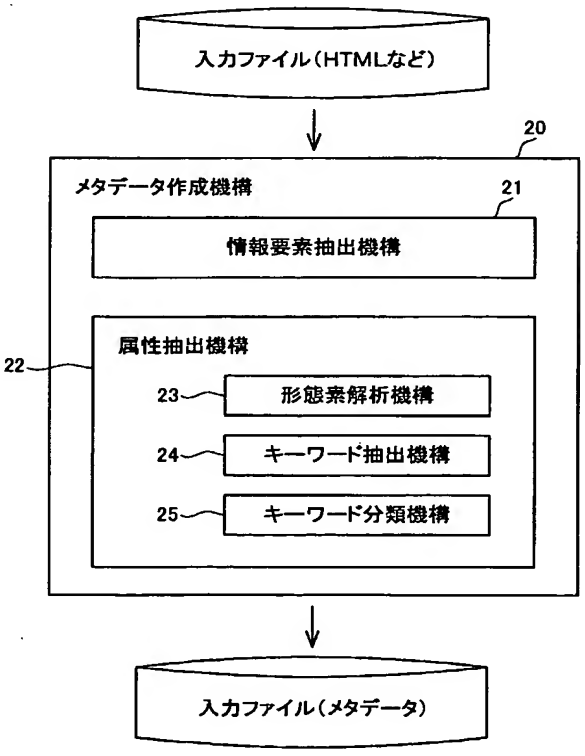


【図３】

登録されたサイトの例

```
<site>
  <url>http://www.ibm.co.jp/</url>
  <name>IBM Japan Homepage</name>
</site>
<site>
  <url>http://www.ibm.com/</url>
  <name>IBM Corporation</name>
</site>
<site>
  <url>http://www.jp.ibm.com/shop/</url>
  <name>日本アイ・ビー・エム[ショッピング]</name>
</site>
<site>
  <url>http://www.jp.ibm.com/developerworks/</url>
  <name>developerWorks</name>
</site>
```

【図４】



【図５】

リンクの例

HTMLファイル中の表現

```
<a href="http://www.ibm.com/jp/software/data/udb/v7/seminar.
html">DB2 UDB V7 無料セミナー開催</a>
```

抽出された情報要素

```
<anchor>
// リンクのタイトル
<DCtitle>DB2 UDB V7 無料セミナー開催</DCtitle>
// URL
<url>http://www.ibm.com/jp/software/data/udb/v7/seminar.html</url>
// 抽出されたキーワード
<kwds>
//          キーワード        // キーワードの分類
<kwd><word>DB</word><class>T0</class></kwd>
<kwd><word>UDB</word><class>T0</class></kwd>
<kwd><word>V</word><class>T0</class></kwd>
<kwd><word>セミナー</word><class>13</class></kwd>
<kwd><word>開催</word><class>13</class></kwd>
<kwd><word>無料</word><class>13</class></kwd>
</kwds>
</anchor>
```

【図６】

テキストブロックの例

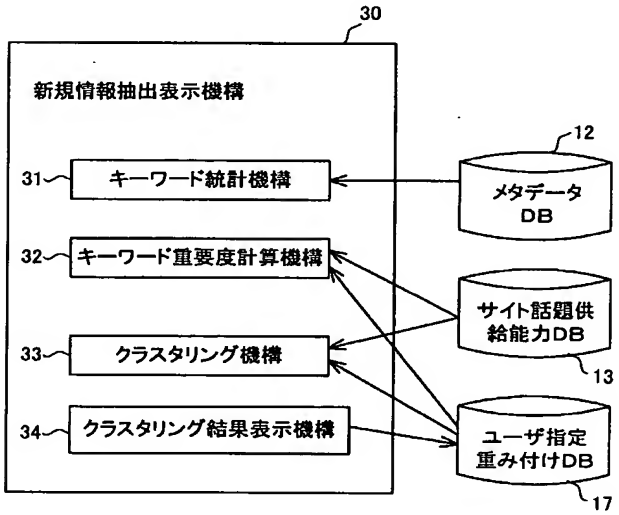
HTMLファイル中の表現

```
インダストリー・トーク / 今週のe-コラム
```

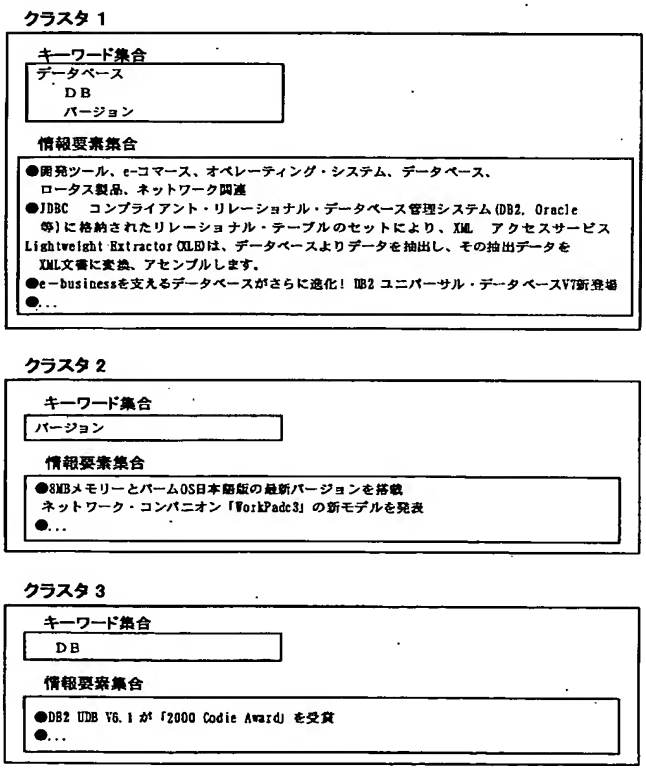
抽出された情報要素

```
<text>
// テキスト部分
<DCdescription>インダストリー・トーク / 今週のe-コラム
</DCdescription>
<kwds>
<kwd><word>e</word><class>T0</class></kwd>
<kwd><word>インダストリー</word><class>T0</class></kwd>
<kwd><word>コラム</word><class>11</class></kwd>
<kwd><word>トーク</word><class>13</class></kwd>
<kwd><word>今週</word><class>11</class></kwd>
</kwds>
</text>
```

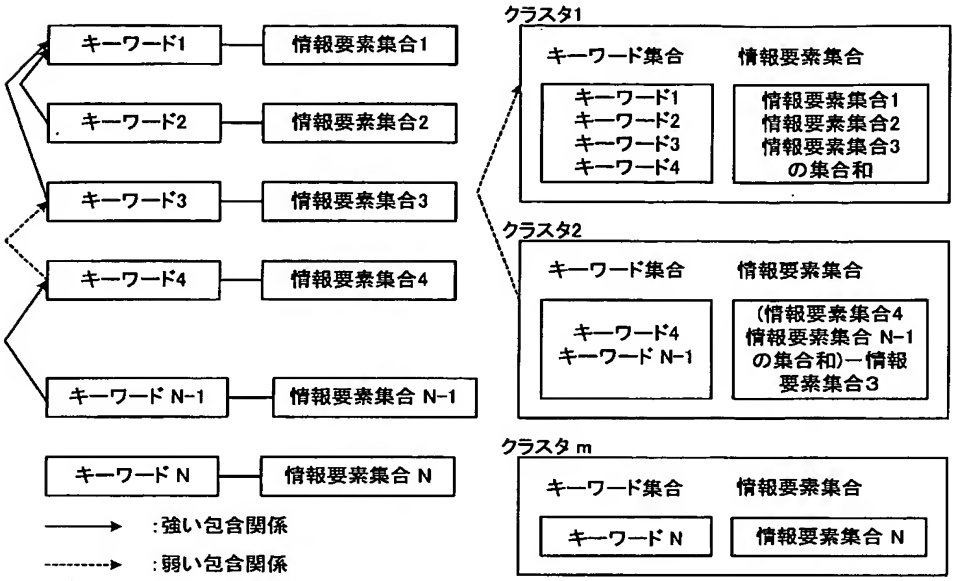
【図７】



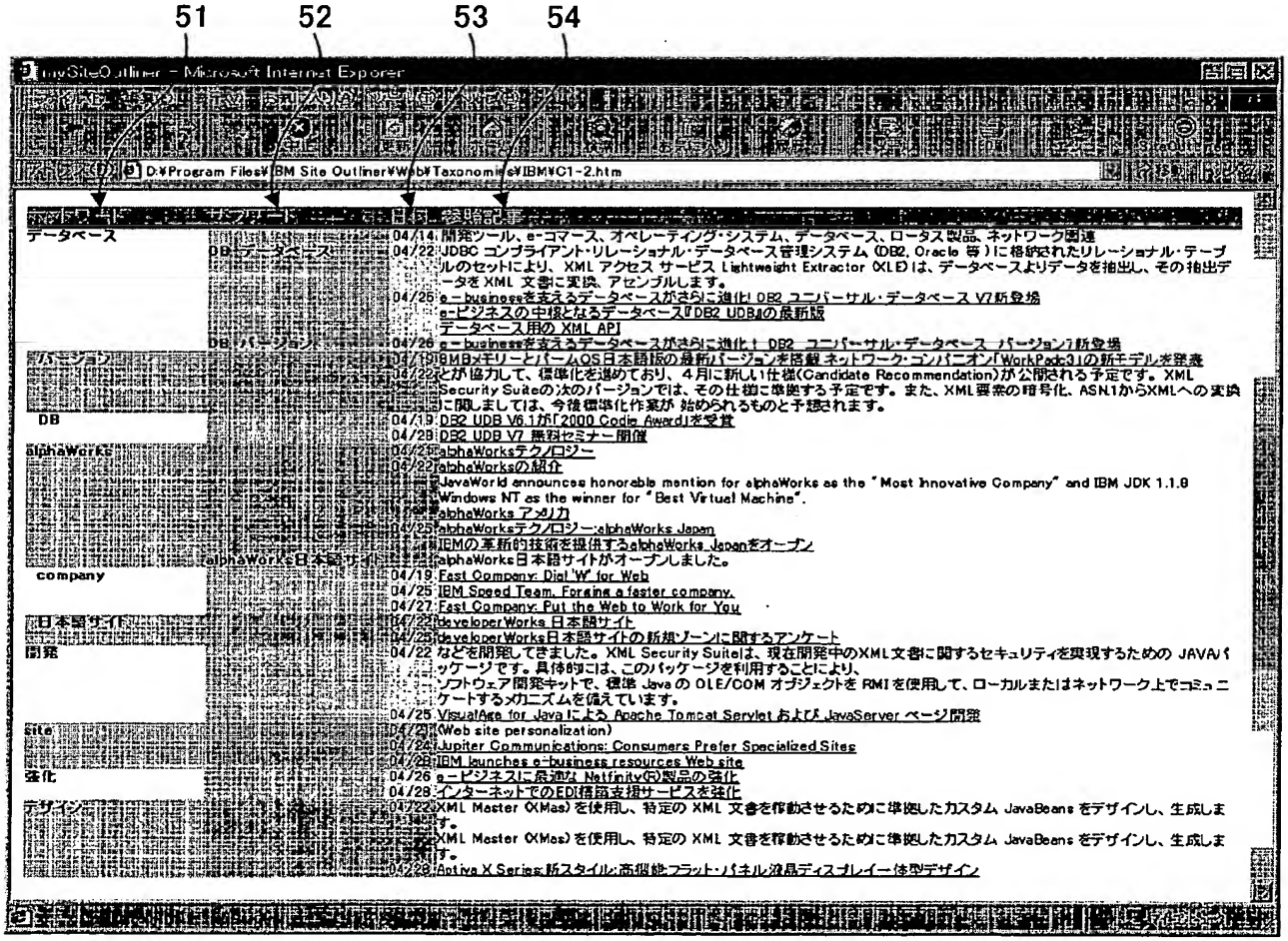
【図１０】



【図９】


キーワード1
情報要素集合1
キーワード2
情報要素集合2
キーワード3
情報要素集合3
キーワード4
情報要素集合4
キーワード N-1
情報要素集合 N-1
キーワード N
情報要素集合 N
:強い包含関係
:弱い包含関係
クラスタ1
キーワード集合
情報要素集合
キーワード1
キーワード2
キーワード3
キーワード4
情報要素集合1
情報要素集合2
情報要素集合3
の集合和
クラスタ2
キーワード集合
情報要素集合
キーワード4
キーワード N-1
(情報要素集合4
情報要素集合 N-1
の集合和)－情報
要素集合3
クラスタ m
キーワード集合
情報要素集合
キーワード N
情報要素集合 N

【図１１】


51
52
53
54
データベース
DB
alphaWorks
company
日本語サイト
開発
Site
強化
デザイン
04/14 開発ツール、e-コマース、オペレーティング・システム、データベース、ロータス製品、ネットワーク関連
04/22 JDBC コンプライアント・リレーショナル・データベース管理システム (DB2, Oracle 等) に格納されたリレーショナル・テーブルのセットにより、XML アクセス サービス Lightweight Extractor (XLE) は、データベースよりデータを抽出し、その抽出データを XML 文書に変換、アセンブルします。
04/25 e-businessを支えるデータベースがさらに進化! DB2 ユニバーサル・データベース V7新登場
e-ビジネスの中核となるデータベース「DB2 UDB」の最新版
データベース用の XML API
04/26 e-businessを支えるデータベースがさらに進化! DB2 ユニバーサル・データベース バージョン7新登場
04/19 IBMメモリーとパームOS日本語版の最新バージョンを搭載 ネットワーク・コンパニオン「WorkPad c3」の新モデルを発表
04/22 とが協力して、標準化を進めており、4月に新しい仕様(Candidate Recommendation)が公開される予定です。XML Security Suiteの次のバージョンでは、その仕様に準拠する予定です。また、XML要素の暗号化、ASN.1からXMLへの変換に関しましては、今後標準化作業が 始められるものと予想されます。
04/19 DB2 UDB V6.1が「2000 Codie Award」を受賞
04/28 DB2 UDB V7 無料セミナー開催
04/22 alphaWorksテクノロジー
04/22 alphaWorksの紹介
JavaWorld announces honorable mention for alphaWorks as the "Most Innovative Company" and IBM JDK 1.1.8 Windows NT as the winner for "Best Virtual Machine".
alphaWorks アメリカ
04/25 alphaWorksテクノロジー:alphaWorks Japan
IBMの革新的技術を提供するalphaWorks Japanをオープン
alphaWorks日本語サイトがオープンしました。
04/19 Fast Company: Dial 'W' for Web
04/25 IBM Speed Team. Forging a faster company.
04/27 Fast Company: Put the Web to Work for You
04/22 developerWorks 日本語サイト
04/25 developerWorks日本語サイトの新規ゾーンに関するアンケート
04/22 などを開発してきました。XML Security Suiteは、現在開発中のXML文書に関するセキュリティを実現するための JAVAパッケージです。具体的には、このパッケージを利用することにより、ソフトウェア開発キットで、標準 Java の OLE/COM オブジェクトを RMI を使用して、ローカルまたはリモートでコミュニケートするメカニズムを備えています。
04/25 VisualAge for Java による Apache Tomcat Servlet および JavaServer ページ開発
04/28 (Web site personalization)
04/24 Jupiter Communications: Consumers Prefer Specialized Sites
04/28 IBM launches e-business resources Web site
04/26 e-ビジネスに最適な Netfinity(R)製品の強化
04/28 インターネットでのEDI構築支援サービスを強化
04/22 XML Master (Mas) を使用し、特定の XML 文書を稼動させるために準拠したカスタム JavaBeans をデザインし、生成します。
XML Master (Mas) を使用し、特定の XML 文書を稼動させるために準拠したカスタム JavaBeans をデザインし、生成します。
04/28 Aptiva X Series:新スタイル:高機能フラット・パネル液晶ディスプレイー体型デザイン

フロントページの続き

(72)発明者 野美山 浩
神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 東京基礎研究所内

Fターム(参考) 5B075 NK31 NR03 NR12 PQ02 PQ34 PR08